PROBLEMAS DE ESTADÍSTICA.

TEMA 1.

1.- Los diámetros interiores de las arandelas producidas por presa pueden medirse con una aproximación de milésimas de pulgas. Las marcas de clase de una distribución de frecuencias de muestras vienen dadas en pulgadas por los nº: 0'321; 0'324; 0'327; 0'333 y 0'336.

Hallar: a) El tamaño del intérvalo de clase.
b) Los límites reales de clase.
c) Los límites de clase.

a) Hallo el tamaño del intérvalo de clase:

$0'324 - 0'321 = 0'003$; $0'003 + 0'001 = 0'004$.

El tamaño es 4.

b) Hallo los límites reales de clase:

x: límite inferior de clase.

y: límite superior de clase.

$$\left.\begin{array}{l} y - x = 0'003 \\ \frac{x+y}{2} = 0'321 \end{array}\right\} \begin{array}{l} y = 0'003 + x \ ; \ y = 0'322 \\ \frac{x + 0'003 + x}{2} = 0'321 ; \ 2x + 0'003 = 0'642 \ ; \ 2x = 0'639 \ ; \ x = 0'319 \end{array}$$

Los límites reales de clase serán: 0'3185 - 0'3225
0'3225 - 0'3255
0'3255 - 0'3285
0'3285 - 0'3315
0'3315 - 0'3345
0'3345 - 0'3375

c) Hallo los límites de clase:

Los límites de clase serán: 0'319 - 0'322 ; 0'322 - 0'325 ; 0'325 - 0'328 ; 0'328 - 0'331 ; 0'331 - 0'334 ; 0'334 - 0'337.

2.- Calcular la proporción de solteros, casados y viudos dentro de un grupo de 150 adultos, conociendo el nº de personas y la proporción de solteros, casados y viudos dentro de cada uno de los subgrupos en los que se divide el grupo 1º, de acuerdo con el cuadro siguiente:

	G.1; n=60	G.2; n=90
solteros	0'30	0'25
casados	0'45	0'55
viudos	0'25	0'20

a) Hallo la proporción de solteros, casados y viudos dentro de l grupo de 150 adultos:

1º Hallo el nº de personas que corresponde al porcentaje dado:

G.1; n=60

solteros: 0'30 => $n_{s_1} = 60 \cdot 30 / 100$; $n_{s1} = 18$ solteros

casados: 0'45 => $n_{c_1} = 60 \cdot 45 / 100$; $n_{c1} = 27$ casados

viudos: 0'25 => $n_{v_1} = 60 \cdot 25 / 100$; $n_{v1} = 15$ viudos

G.2; n=90

solteros: 0'25 => $n_{s2} = 90 \cdot 25 / 100$; $n_{s2} = 22'5$ solteros

casados: 0'55 => $n_{c2} = 90 \cdot 55 / 100$; $n_{c2} = 49'5$ casados.

viudos: 0'20 => $n_{v2} = 90 \cdot 20 / 100$; $n_{v2} = 18$ viudos

$\frac{n_{s1} + n_{s2}}{n_{TP}} = \frac{18 + 22'5}{150}$ => Porcentaje de solteros: 0'27.

$\frac{n_{c1} + n_{c2}}{n_{TP}} = \frac{27 + 49'5}{150}$ => Porcentaje de casados: 0'51.

$\frac{n_{v1} + n_{v2}}{n_{TP}} = \frac{15 + 18}{150}$ => Porcentaje de viudos: 0'22.

3.- El valor menor de una variable observada es 21'2, y el mayor valor es 28'3. Establecer las clases y marcas de clases correspondientes.

a) Establezco las clases y marcas de clases correspondientes:

1º Hallamos el nº de valores ≠ que puede tomar la función:

* Tú diferencias entre el 28'3 y el 21'2 y no entre el 28 y el 21.

28'3 - 21'2 = 7'1

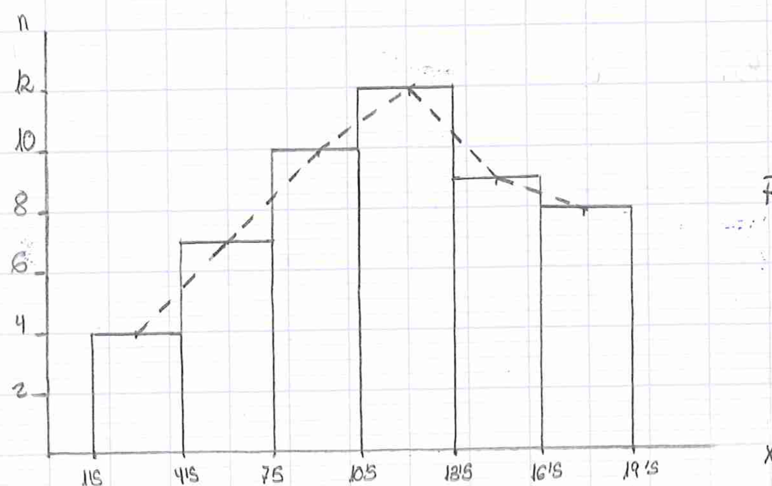
7'1 + 0'1 = 7'2 => 72 valores ≠ toma la variable.

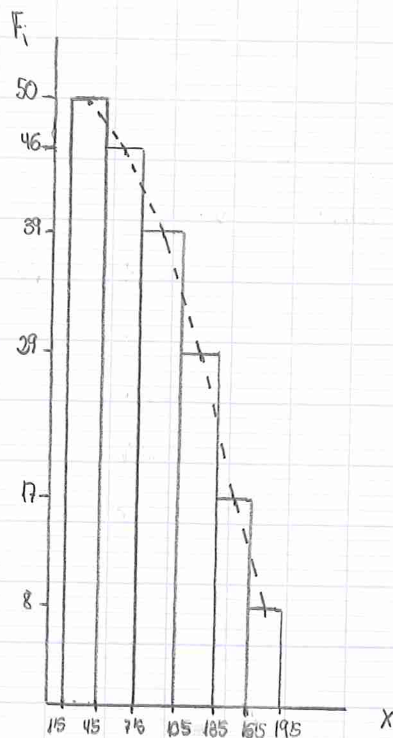
CLASES		MARCA	CLASES		MARCA
21'2 - 21'9	→	21'5	24'7 - 25'4	→	25
21'9 - 22'6	→	22'2	25'4 - 26'1	→	25'7
22'6 - 23'3	→	22'9	26'1 - 26'8	→	26'4
23'3 - 24	→	23'6	26'8 - 27'5	→	27'1
24 - 24'7	→	24'3	27'5 - 28'3	→	27'9

4.- A partir de la siguiente distribución de frecuencias, calcular frecuencias acumuladas y las proporciones y porcentajes sin acumular y acumulados. Dibujar además, el histograma y el polígono de frecuencias, considerando éstas sin acumular y acumuladas.

x	n	Fi	fi relativa	fi %	Fi relativa	Fi %
17-19	8	8	8/50 = 0'16	16%	8/50 = 0'16	16%
14-16	9	17	9/50 = 0'18	18%	17/50 = 0'34	34%
11-13	12	29	12/50 = 0'24	24%	29/50 = 0'58	58%
8-10	10	39	10/50 = 0'2	20%	39/50 = 0'78	78%
5-7	7	46	7/50 = 0'14	14%	46/50 = 0'92	92%
2-4	4	50	4/50 = 0'08	8%	50/50 = 1'00	100%

Gráficos:





FRECUENCIA RELATIVA.

5.- Calcular la media de los siguientes valores agrupándolos primero por intervalos de amplitud igual a 5 y después por intervalos de amplitud igual a 10. Estos valores son: 49, 48, 43, 42, 49, 41, 42, 43, 43, 44, 44, 51, 53, 54, 51, 59, 58, 57, 56, 54, 51, 54, 53, 64, 62, 64, 63, 62, 61, 62, 68, 68, 67, 66, 69.

VALORES	FRECUENCIA	VALORES	FRECUENCIA	VALORES	FRECUENCIA
41	1	¿? 55	0	67	1
42	2	56	1	68	2
43	3	57	1	69	1
44	2	58	1		
¿? → 45, 46, 47	0	59	1		
48	1	¿? 60	0		
49	2	61	1		
¿? 50	0	62	3		
51	3	63	1		
¿? 52	0	64	2		
53	2	¿? 65	0		
54	3	66	1		

a) Calculo la media de los valores dados:

1. Agrupándolos en intervalos de amplitud igual a 5:

x	n	x_i
40-44	8	42
46-49	3	47
50-54	8	52
55-59	4	57
60-64	7	62
65-69	5	67

* Tanto el valor del límite inferior del intervalo como el del superior están incluidos.

– Se trata de una estadística del tipo III:

Hallo la media aritmética:

$$\bar{X} = \frac{\sum x_i \cdot f_i}{N} ; \bar{X} = \frac{42.8+47.3+52.8+57.4+62.7+67.5}{35} = \frac{1890}{35} ; \bar{X} = 54$$

6. – Argumentar la posibilidad de obtener la moda en las siguientes distribuciones:

X	%
menos de 25	5
25-29	10
30-34	17
35-39	28
40-49	18
50-59	12
60-69	6
más de 70	4
	100

x	%
menos de 1000	34
1000-2000	25
2000-5000	20
5000-8000	12
8000-11000	5
más de 11000	4
	100

Hallo la moda en ambos casos:

En el primer caso nos encontramos ante una estadística del tipo III. La amplitud de los intervalos no es constante. El intervalo modal será

5.

el de 35-39 (f_i = 28). Una vez encontrado el intervalo modal existen 2 métodos para la obtención de la moda:

A.- Tenemos como moda la marca de clase del intervalo modal ⇒

$Mo = 35+39/2$; $\underline{Mo = 37}$

B.- Por la fórmula ⇒

$$Mo = Linf_j + \frac{f_{i+1}}{f_{i-1} + f_{i+1}} \cdot h_j$$

$$Mo = 34'5 + \frac{18}{17+18} \cdot 5; \quad \underline{Mo = 37}$$

No se puede pues el límite inferior no está definido. clase abierte.

→ En el segundo caso nos encontramos ante una estadística de tipo III: La amplitud de los intervalos no es constante. El intervalo será el de menos de 1000 (f_i = 34). Una vez hallado el intervalo modal existen 2 métodos para la obtención de la moda. Sin embargo hemos de tener en cuenta que el intervalo modal carece de límite inferior. (Tomaremos como límite inferior el cero):

A.- Tenemos como moda la marca de clase del intervalo modal ⇒

$Mo = 0+1000/2$; $\underline{Mo = 500}$.

B.- Por la fórmula ⇒

$$Mo = Linf_j + \frac{f_{i+1}}{f_{i-1} + f_{i+1}} \cdot h_j$$

$$Mo = -0'5 + \frac{25}{25+0} \cdot 1000$$

7.- La tabla siguiente nos muestra la distribución de las longitudes de una muestra de insecto en estudio, medido en milímetros.

Longitudes en mm.	Frecuencia	Fi
9'3 - 9'7	2	2
9'8 - 10'2	5	7
10'3 - 10'7	12	19
10'8 - 11'2	17	36
11'3 - 11'7	14	50

11'8 - 12'2	6	56
12'3 - 12'4	3	59
12'8 - 13 - 2	1	60
	60	

Se pide: a) Calcular la media, mediana, y moda.
b) Calcular los cuartiles 1º, 2º y 3º.

a) Calculo la media, mediana, y moda. (Se trata de 1 estadística del tipo III).

1. Media → $\bar{X} = \sum x_i f_i / N$.

$$\bar{X} = \frac{2.9'5 + 5.10 + 12.10'5 + 17.11 + 14.11'5 + 6.12 + 3.12'5 + 1.13}{60} = \frac{665'5}{60}; \bar{X} = 11$$

La media vale 11.

2. Mediana → Hallamos la mediana por el 1er método: (fórmula).
$N/2 = 30$ => El intervalo crítico será el 10'8 - 11'2

$$Me = Linf_i + \frac{N/2 - F_{i-1}}{f_i} \cdot l_i \qquad Me = 10'75 + \frac{30 - 19}{17} \cdot 0'5; \ Me = 11$$

La mediana vale 11.

3. Moda → Hallamos la moda por el 1er método:
El intervalo modal será el 10'8 - 11'2 => La marca de clase será: $\frac{10'8 + 11'2}{2} = 11$.
La moda vale 11.

b) Calculo los cuartiles 1º, 2º y 3º: lo hacemos mediante fórmula:

1. Cuartil 1 => $Q_1 = 10'75 + \frac{1 \cdot 60/4 - 19}{17} \cdot 0'5$; $Q_1 = 10'6$

2. Cuartil 2 => $Q_2 = 10'75 + \frac{2 \cdot 60/4 - 19}{17} \cdot 0'5$; $Q_2 = 11'1$

3. Cuartil 3 ⇒ $Q_3 = 10'75 + \frac{3 \cdot 60/4 - 19}{17} \cdot 0'5$; $Q_3 = 11'5$

8.– Dados los siguientes datos no agrupados: 1, 9, 3, 7, 8, 8. Calcular:

a) La varianza.

b) La desviación típica.

c) La desviación media.

9.– Calcular los mismos apartados que en el problema nº 8 para la siguiente agrupación de datos:

x	n
75-78	2
71-74	3
67-70	8
63-66	7
59-62	4
55-58	1

10.- Las puntuaciones en una prueba de inteligencia abstracta han sido las siguientes: 91 92 83 81 88 94 91 87 90 94 85 85 93 90 89 86 87 89 85 89.

Preparar una distribución de frecuencias sin acumular y acumuladas, introduciendo intervalos de amplitud 4. Calcular las proporciones y porcentajes sin acumular y acumulados.

81 → 1	84 → 0	87 → 2	90 → 2	93 → 1	96 → 0
82 → 0	85 → 3	88 → 1	91 → 2	94 → 2	
83 → 1	86 → 1	89 → 3	92 → 1	95 → 0	

x	fi	Fi	fi relativa	fi %	Fi relativa	Fi %
81-84	2	2	2/20=0'10	10	2/20=0'10	10
85-88	7	9	7/20=0'35	35	9/20=0'45	45
89-92	8	17	8/20=0'40	40	17/20=0'85	85
93-96	3	20	3/20=0'15	15	20/20=1	100
	20					

11.- Los momentos de 2º orden con respecto a la media de dos distribuciones, son 9 y 11, mientras que el momento de 3º orden con respecto a la media son -9'1 y -12'3 respectivamente. ¿Qué distribución está más sesgada a la izquierda?

NOTA: el momento de 2º orden con respecto a la media es la varianza.

12.- De un informe del INE, hemos obtenido los datos correspondientes a la distribución porcentual de hogares según el nivel de ingresos, ordenándolos en la siguiente tabla:

Intervalos de renta	Porcentaje de hogares	
(en miles de pesetas)	Conjunto rural	Conjunto Nacional
hasta 69	6'30	7'58
69 - 84	9'83	6'56
84 - 120	20'87	12'22
120 - 180	27'63	21'92
180 - 240	15'00	18'48
240 - 480	18'27	27'10
480 - 700	1'65	4'40
más de 700	0'45	1'94.

NOTA: El límite superior del intervalo, no entra en dicho intervalo.

Del mismo informe del INE, hemos extraído y ordenado los porcentajes correspondientes a los deciles de hogares según se trate de municipios de

menos de 2000 habitante ó de más de 50000 obteniendo esta tabla:

Deciles de lugares	Conjunto de lugares	
	menos de 2000 hab.	más de 50000 hab.
1	1'85	2'39
2	3'46	4'33
3	4'90	5'63
4	6'35	6'74
5	7'73	7'80
6	9'19	8'96
7	10'83	10'36
8	13'90	12'18
9	16'19	15'21
10	26'60	26'42

Se pide: Hacer un análisis del reparto de ingresos en la España rural.

En la otra cara de la hoja.

13.- Tres profesores de economía registraron una calificación media en sus exámenes 79; 82 y 84. Sus clases estaban formadas por 32; 25 y 17 estudiantes respectivamente.

Determinar la calificación media para todas las clases.

Clase A: 32 alumnos, nota media → 79.

Clase B: 25 alumnos; nota media → 82.

Clase C: 17 alumnos; nota media → 84.

Determino la nota media para todas las clases:

$$N.M = \frac{79.32 + 82.25 + 84 \cdot 17}{32 + 25 + 17} = \frac{2528 + 2050 + 1428}{74} = \frac{6006}{74} = 81.$$

<u>La nota media para las tres clases es de 81</u>

⇒ (12)

TABLA 1 < Ⓐ Interv. de rentas Porcentaje de hogares. conjunto rural.
Ⓑ Interv. de rentas Porcentaje de hogares: conjunto nacional

No podemos calcular la media porque los intervalos no están definidos ya que no podemos conocer las marcas de clases.

Hay calcular la mediana y la moda.

Calculamos la asimetría para ver si está sesgada a la izq o a la dcha.

Calcular la concentración: en este problema no podemos calcular las intervalo - hasta 69 - y - más de 700 - Índice de Lorentz o de Gini.

TABLA 2 < Ⓐ Deciles de hogar → Porcentaje de ingresos en pob. de - de 2000 hab.
Ⓑ Deciles de hogar → Porcentaje de ingresos en pob. de + de 50000 hab.

Calcular la concentración, Identificar las "p" y "Q".

(TABLA 1.A:)

Análisis: Comparar datos conj. rural y conj. nacional.

Medida de posición que podemos obtener: moda, mediana y media.

Media → No porque no hay intervalos definidos.

Mediana → Buscar intervalo crítico, 120 - 180, aplicando la fórmula nos dará (p.e: 140), donde el 50% de los hogares obtienen una renta inferior a 140 y otro 50% una renta superior a 140.

En el problema nos dan como datos las frecuencias relativas, como para la fórmula necesitamos frecuencias absolutas, tomamos 100 como total.

Moda → Intervalo modal (120 - 180). Cogemos "150" (marca de clase) del intervalo modal.

(TABLA 1.B:)

Intervalo crítico: (contar todas las frecuencias y la que pasa de 50, ese es el intervalo crítico) 180-240 → intervalo crítico.

Moda → intervalo modal.

Conclusión: las rentas más repetidas son de 150 en conj. rural, en el conj. nacional serán de 360. (marca de clase del intervalo modal).

Podríamos hallar también la dispersión, aunque no es muy útil, tomando la mediana. (porque no podemos calcular la media).

Podríamos obtener la varianza (tipo de dispersión)

" " una medida de asimetría: para hallar el sesgo hacia la izq. → intervalo de renta más bajo con mayor frecuencia; hacia la drcha → intervalo de mayor renta con la frecuencia más alta. (E. → la mayoría de los hogares tienen renta menor o (drcha) la mayoría de los hogares tienen un nivel de ingresos alto).

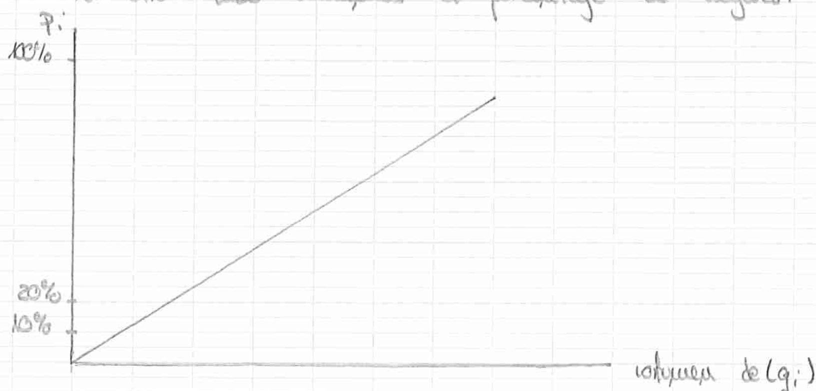
Podríamos hallar la concentración (que se halla en todos los ejercicios de niveles de rentas). Para.

q_i: frec. acumuladas en porcentaje de volumen de (cant. tot. de ingresos: frec. x cant. de renta).

En este caso: marcas de clase de cada intervalo x su frecuencia = volumenes de.

p_i: Hallamos el porcentaje, lo acumulamos.

Por otro lado hallamos el porcentaje de hogares:



Si el índice de Gini se acerca a ∅: el reparto es eq.

Si se acerca a 1 hay mucha concentración (mal reparto).

(TABLA 2.A:)

Decil	P
1	1'85
2	3'46
3	4'90

→ Quiere decir que el 10% de los lugares con – renta tiene un 1'85% del total de ingresos.

→ Que el 20% de lugares con – ingresos posee el (1'85+3'46)% del total de ingresos (porque el 20% es acumulado)

→ El 30% de lugares con – ingresos posee (1'85+3'46+4'9)% de los ingresos totales.

¿Qué nivel de renta tiene la población con menos de 1000?

No podemos calcular ni "media" ni "mediana" porque no tenemos ningún nivel de ingresos. Lo mismo pasa con la moda y la medida de dispersión.

Tampoco podríamos hallar la asimetría basada en el método de los cuartiles, medimos distancias de abscisas, en este caso con los niveles de ingresos, que no podemos obtener. De ahí que este método no sirve.

Lo único que podemos hallar es la concentración: si los ingresos están + concentrados en pob. de < de 2000 y pob. de + de 50000. Si tenemos que calcular la concentración tendremos que calcular los "pi" y "qi".

Pob < 2000

p_i	q_i
10	1'85
20	5'31 (1'85+3'46)
30	10'21
40	16'56
50	24'29
60	33'18
70	44'31
80	57'21
90	70'11
100	86'3

Pob > 50000

p_i	q_i
10	
20	
30	
40	
50	
60	
70	
80	
90	
100	

p_i: frecuencias en porcentajes y acumulados

q_i: "volumen de" en porcentaje y acumulados.

p_i / q_i (ejes de gráfico vacío)

p_i / q_i (ejes de gráfico vacío)

¿? => 14. _ De la siguiente distribución de frecuencias acumuladas en sentido decreciente:

VALORES	frecuencias
más de -10	22
" " 10	19
" " 20	11
" " 30	2

Deducir la distribución acumulada creciente, sabiendo que como máximo $x = 150$.

VALORES	frecuencias.
-10 o menor	
10 o menor	
20 o menor	
30 o menor	

15._ La distribución del nº de niñas por familia en un colegio, es la siguiente:

Nº de niñas:	8	7	6	5	4	3	2	1	0
Familias:	1	2	6	8	20	38	60	60	35.

Se pide: a) Calcular el número medio de niñas por familia.

b) Hallar los cuartiles 1º y 3º de estos datos.

a) Calculo el número medio de niñas por familia:

$$X = \frac{8.1+7.2+6.6+5.8+4.20+3.38+2.60+1.60+0.35}{1+2+6+8+20+38+60+60+35} = \frac{8+14+36+40+80+114+120+60+0}{230}$$

$X = 472/230 = 2 \Rightarrow$ El número medio de niñas por familia es de 2.

b) Calculo los cuartiles 1º y 3º de estos datos.

Cuartil 1º: Observación 57'5 -> nº de niñas: 1.

Cuartil 3º: Observación 172'5 -> nº de niñas: 3.

16.- El profesor de inglés de los alumnos de segundo año de un colegio propone un test de gramática a la clase y obtiene el siguiente resultado:
37, 35, 43, 45, 45, 35, 35, 42, 40, 46, 35, 34, 44, 43, 34, 40, 38, 37, 40, 38, 36.

Hallar la desviación típica del grupo, e interpretar el resultado.

17.- Decir si pueden calcularse la media aritmética, la mediana y la moda de la siguiente distribución de frecuencias.

intervalos	frecuencia
menos de 10	21
de 10 a 20	62
de 20 a 40	275
de 40 a 80	130
más de 80	26

18.- Hallar la frecuencia correspondiente al tercer intervalo de la siguiente distribución, sabiendo que su media aritmética es igual a 11'5.

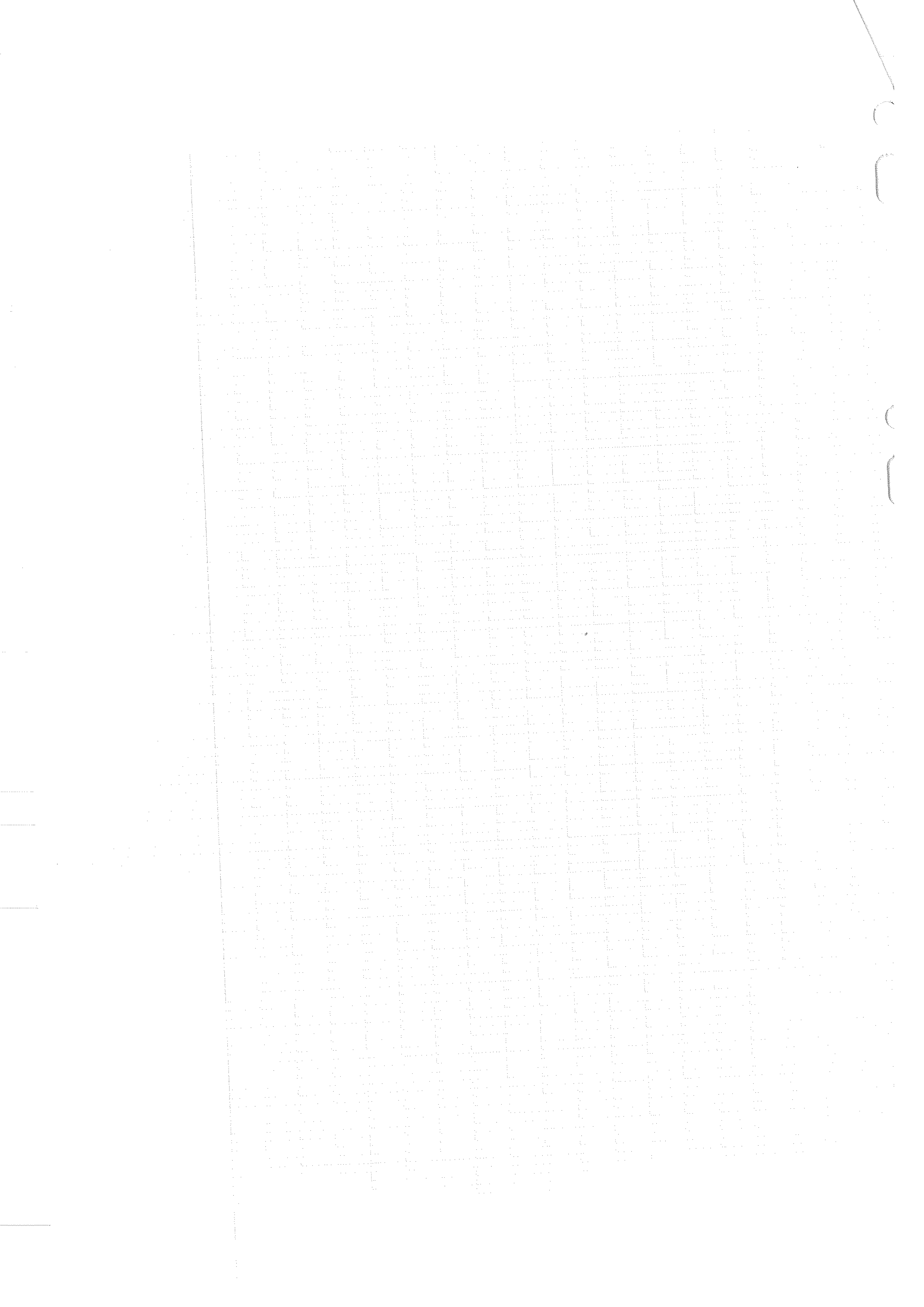
intervalos	frecuencias
4 - 6	4
6 - 10	5
10 - 16	-
16 - 20	3
20 - 30	1

Nota: El límite superior de cada intervalo, no entra en dicho intervalo.

a) Hallo la frecuencia correspondiente al 3º intervalo:

$$11'5 = \frac{2 \cdot 4 + 4 \cdot 5 + 6 \cdot x + 4 \cdot 3 + 10 \cdot 1}{13 + x} ;\ 149'5 + 11'5x = 8 + 20 + 6x + 12 + 10 ;\ 149'5 + 11'5x = 50 + 6x$$

19.- En una distribución de frecuencias, la diferencia entre el 1º cuartil y la mediana es doble de la que existe entre la mediana y el 3º cuartil, y la diferencia entre el primer valor de la variable y el 1º cuartil es igual a la existencia entre el 3º cuartil y el último valor de la variable. Razonar cuál será el signo de la asimetría de dicha distribución.



Dada una sociedad con los siguientes datos. (S. A).

Capital invertido (en millones) (VARIABLE)	Nº accionistas (FRECUENCIA)	F_i (frec. acumulada)
10	100	100
20	56	156
30	81	237
(40)	495	732
50	80	812
90	40	852
100	75	927
	927	

a) Calcúlese la mediana (un valor de la variable que está justo en la mitad).

- Estadística TIPO II
- 'n' es 1 nº impar.
- $N/2 = 927/2 = 463'5$
- El valor inmediata/ superior al 463'5 en la frecuencia acumulada (40)
- $n+1 = 40+40/2$

$$Me = \frac{\overset{463}{n} + \overset{464}{(n+1)}}{2} \qquad N = 927 = (2n+1) = 2(463)+1.$$

$$\underline{\underline{Me = 40}} \Longleftarrow Me = \frac{40+40}{2}$$

IMPRESCINDIBLE ¿Cómo se interpreta ésto?

El 50% de los accionistas tendrán invertido un capital ≤ 40 millones ó que hay 463 accionistas cuya inversión será ≤ 40 millones.

Pandolfo ha decidido convertirse en accionista del club local, q recientemente ha decidido transformarse en S.A con 1 capital de 100 millones. Una vez finalizado el proceso el club le ha enviado 1 relación de accionistas así como el capital desembolsado por cada uno. Tras 1 paciente recuento Dandolfo ha llegado a tener la siguiente tabla.

CAPITAL INVERTIDO (mils)	Nº ACCIONISTAS
10	100
20	58
30	81
40	495
50	80
70	40
90	40
100	75
150	20
200	20
300	15
500	18
750	15
1000	20
3000	2

A la vista de esta tabla se plantea las cuestiones: ¿Cuál es la inversión media por accionista? ¿El promedio calculado es representativo?

En la Junta de accionistas tendrán derecho a voto quienes hallan desembolsado al - 40000 pts. ¿Qué proporción de accionistas queda excluida?

¿Existe mucha concentración en el reparto del capital del club